



NATIONAL INSTITUTE OF TECHNOLOGY PATNA
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

MID Semester Examination Jan - June 2022

B. Tech: CSE -VIth semester

Course Name: Data Mining & Warehousing

Course Code: **CS6403**

Max. Marks : 30

Maximum Time: 2 Hours

Instructions:

1. Attempt all questions.
2. Assume any suitable data, if necessary.
3. **Write your roll number, name and subject code on each and every page top-left.**
4. Answer all parts of the question at the same place.
5. **Upload the answer scripts within the given deadline.**

Q.No	Question	Marks	CO	BL																		
1. a. b.	What is supervised and semi-supervised learning? How do we handle missing values in the data cleaning process?	3+3	CO1	L1																		
2. a. b.	Explain various techniques involved in the data transformation? Normalize the following group of data: 200, 300, 400, 600, 1000 using <i>min-max normalization</i> by setting min = 0 and max = 1	3+3	CO1	L2, L3																		
3.	What do you mean by the OLAP? Explain various OLAP operations in multidimensional data model?	6	CO2	L1, L2																		
4.	<table border="1"><tr><td>A</td><td>35</td><td>49</td><td>25</td><td>33</td><td>65</td><td>25</td><td>45</td><td>51</td></tr><tr><td>B</td><td>8</td><td>9</td><td>5</td><td>9</td><td>11</td><td>5</td><td>11</td><td>12</td></tr></table> <p>Identify the correlation between above two numerical attributes A and B? (<i>It is compulsory to show all the intermediate steps in Pearson correlation.</i>)</p>	A	35	49	25	33	65	25	45	51	B	8	9	5	9	11	5	11	12	6M	CO1	L3
A	35	49	25	33	65	25	45	51														
B	8	9	5	9	11	5	11	12														
5.	<p>An age group of 1550 people was surveyed. The gender of each person was male or female. Each person preferred reading material was either fiction or non-fiction. The contingency table of this scenario is given below:</p> <table border="1"><tr><td></td><td>Male</td><td>Female</td></tr><tr><td>Fiction</td><td>300</td><td>250</td></tr><tr><td>Non-fiction</td><td>100</td><td>900</td></tr></table> <p>Find the χ^2 (Chi-square) value between the attribute <i>gender</i> and <i>preferred reading</i>. (<i>It is compulsory to show all the intermediate steps in the solution.</i>)</p>		Male	Female	Fiction	300	250	Non-fiction	100	900	6M	CO1	L1, L2									
	Male	Female																				
Fiction	300	250																				
Non-fiction	100	900																				

*****ALL THE BEST*****